A

VENABLE, BAETJER, HOWARD & CIVILETTI, LLP
Including professional corporations

1201 New York Avenue, N.W., Suite 1000
Washington, D.C. 20005-3917
(202) 962-4800, Fax (202) 962-8300
www venable.com

OFFICES IN

WASHINGTON, D.C.
MARYLAND
VIRGINIA

VENABLE
ATTORNEYS AT LAW

NONPROVISIONAL PATENT APPLICATION TRANSMITTAL RULE §1.53(b) IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Venable
P.O. Box 34385
Washington, DC 20043-9998
Telephone: (202) 962-4800
Facsimile: (202) 962-8300

Docket No. 32011-167412

Date: October 27, 2000

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Transmitted herewith for filing under 37 C.F.R. §1.53(b) is a nonprovisional patent application:

For (Title): SPEECH SYNTHESIS DEVICE

By (Inventors): Yukio TABEI

☒ 25 pages of Specification/Claims 1-20/Abstract are attached.
☒ Formal drawings (Fig. 1-9; 9 sheets) is attached.
☒ A Declaration and Power of Attorney is filed herewith.
☒ An assignment of the invention to **OKI ELECTRIC INDUSTRY CO., LTD.**, is filed herewith, along with form PTO-1595 and a check for $40.00.
☒ An Information Disclosure Statement is filed herewith, along with form PTO-1449, and 2 reference.
☐ A Statement to establish small entity status under 37 C.F.R. §§1.9 and 1.27 is filed herewith.
☐ A Preliminary Amendment is filed herewith.
☐ Please amend the specification by inserting before the first line the sentence --This nonprovisional application claims the benefit of U.S. Provisional Application No. ___, filed ___.--
☒ Priority of foreign application No. 2000-075831 filed August 4, 2000 in JAPAN is claimed (35 U.S.C. §119).
☒ A certified copy of the above corresponding foreign application is filed herewith.

The filing fee is calculated below and includes claim status after entry of any Preliminary Amendment noted above:

			SMALL ENTITY			LARGE ENTITY	
FOR:	NO. FILED	NO. EXTRA	RATE	FEE	OR	RATE	FEE
BASIC FEE				$ 345	OR		$ 710
TOTAL CLAIMS	6 - 20	= 0	x 9 =	$	OR	x 18	$
INDEP CLAIMS	1 - 3	= 0	x 39 =	$	OR	x 78	$
☐ MULTIPLE DEPENDENT CLAIMS			+130 =	$	OR	+260	$
			TOTAL	$	OR	TOTAL	$ 710

☒ A check in the amount of **$710.00** is attached hereto. Except as otherwise noted herein, the Commissioner is hereby authorized to charge any other fees that may be required to complete this filing, or to credit any overpayment, to Deposit Account No. 22-0261.

☐ This application is entitled to small entity status. DO NOT charge large entity fees to our Deposit Account.

Respectfully submitted,

James R. Burdett
Registration No. 31,594

#248039

SPEECH SYNTHESIS DEVICE

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to a rule-based speech synthesis device that synthesizes speech, and more particularly to a rule-based speech synthesis device that synthesizes speech from an arbitrary vocabulary.

Description of Related Art

Text-to-speech conversion (the conversion of a text document into audible speech) has hitherto been configured from a text analysis part and a rule-based speech synthesis part (parameter generation part and waveform synthesis part).

Text containing a mixture of *kanji* and *kana* characters (a Japanese-language text document) is input to the text analysis part, where this document is subjected to morphological analysis by referring to a word dictionary, the pronunciation, accentuation and intonation of each morpheme are analyzed (if necessary, syntactic and semantic analysis and the like are also performed), and then phonological symbols (intermediate language) with associated prosodic symbols are output for each morpheme.

In the parameter generation part, prosodic parameters such as pitch frequency patterns, phoneme duration times, pauses and amplitudes are set for each morpheme.

In the waveform synthesis part, speech synthesis units in the target phoneme sequence (intermediate language) are selected from previously stored speech data, and waveform synthesis processing is performed by concatenating/modifying the reference data of these speech synthesis units according to the parameters determined in the parameter generation part. The type of speech synthesis units that have been tried out is phonemes, syllables (CV), and VCV/CVC (C = consonant, V = vowel). Although phonemes have the least number of possible representations, it is essential to incorporate rules for coarticulation, which is not easy to do. Consequently, the resulting synthesized speech has had poor quality, and phonemes are now seldom used as speech synthesis units. On the other hand, CV, VCV and CVC units include coarticulation within each unit. For example, since a VCV type comprises a consonant between two vowels, the consonant part is very clear. And since a CVC type is concatenated with consonants which have small amplitude, the concatenation distortion is small. Recently, units consisting of even larger phonetic chain have also been partially used as speech synthesis units.

As the speech data in the speech synthesis units, a method has come to be used whereby original audio waveforms are used unaltered, and based on this, high quality synthesized sound is obtained with little degradation of quality.

To obtain more natural-sounding synthesized speech with the abovementioned conventional text-to-speech conversion, the way in which the parameters in the abovementioned parameter generation part (pitch frequency pattern, phoneme duration time, pauses, amplitude) are appropriately controlled to approximate natural speech while considering the type of speech synthesis units, the speech segment quality and the synthesis procedure is of great importance.

Of these parameters, methods for controlling the phoneme duration time in particular have hitherto been described in Reference 1 (Japanese Patent Application Laid-Open No. S63-46498) and Reference 2 (Japanese Patent Application Laid-Open No. H4-134499).

The techniques described in the abovementioned References 1 and 2 are methods which use a statistical model (Hayashi's first method of quantification model) to obtain control rules by analyzing a large amount of data. As is well known, a Hayashi's first method of quantification is one of multivariate analysis technique wherein the target external criterion (phoneme duration time) is calculated based on qualitative factors, and is formulated as shown in Formulae (1) through (3) below.

That is, if *j* is the *i*th data element item, *k* is the category to which it belongs, and $x(jk)$ is the category quantity thereof (the coefficient associated with the category), then the estimated values $y(i)$ are given by Formula (1).

$$y(i) = \sum_{j} \sum_{k} x(jk)\ \delta(jk) \qquad \ldots (1)$$

where:

$$\delta(jk) = 1 \text{ (when data } i \text{ corresponds to category } k \text{ of item } j\text{)}$$
$$= 0 \text{ (otherwise)} \qquad \ldots (2)$$

$x(jk)$ is determined by the method of least square. That is, it is determined by minimizing the squared error between the estimated values $y(i)$ and the actual measured values $Y(i)$.

$$\sum_{i} \{y(i) - Y(i)\}^2 \rightarrow \text{minimum} \qquad \ldots (3)$$

The equation has to be solved by partially differentiating Formula (3) by $x(jk)$. When a computer is used to perform real calculations based on Formula (3), it results in a numerical analysis problem to solve simultaneous equations.

In the abovementioned conventional phoneme duration time controling method, categorization into Hayashi's first method of quantification form does not always work well, making it impossible to achieve adequate estimation precision. Also, these conventional methods make no mention of methods for setting the closing length in phonemes having a closing interval (such as unvoiced plosive consonants). Accordingly, there have hitherto been no methods for appropriately controlling the closing interval length, which is of great perceptual importance.

The principal object of the present invention is to provide a rule-based speech synthesis device that can estimate phoneme duration times more accurately and has smaller estimation errors and better control functions, and in particular it aims to provide a suitable closing time length control method for phonemes having a closing interval (such as unvoiced plosive consonants), and as a result, an object of the present invention is to provide a rule-based speech synthesis device with improved quality.

SUMMARY OF THE INVENTION

Consequently, the rule-based speech synthesis device of the present invention is a rule-based speech synthesis device that generates arbitrary speech by selecting previously stored speech synthesis units, concatenating these selected speech synthesis units, and controlling the prosodic information, and which is provided with a phoneme duration time setting means that estimates and controls the closing interval length of phonemes having a closing interval separately from the vowel length and the consonant length.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the present invention will be better understood from following description taken in connection with accompanying drawings, in which:

Figure 1 is a block diagram showing one embodiment of a speech synthesis device (text-to-speech conversion device) relating to this invention;

Figure 2 shows the configuration of the phoneme duration time setting part in a first embodiment of this invention;

Figure 3 shows the configuration of the phoneme duration time setting part in a second embodiment of this invention;

Figure 4 shows the configuration of the phoneme duration time setting part in a third embodiment of this invention;

Figure 5 shows the configuration of the phoneme duration time setting part in a fourth embodiment of this invention;

Figure 6 shows the classes of consonants prefixed by a closing length;

Figure 7 illustrates the operation of the closing length classification part, the closing length learning part and the closing length estimation part in the second embodiment of this invention;

Figure 8 illustrates the operation of the vowel length classification part, the vowel length learning part and the vowel length estimation part in the third embodiment of this invention; and

Figure 9 illustrates the operation of the consonant length classification part, the consonant length learning part and the consonant length estimation part in the third embodiment of this invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Embodiments of the present invention will be described in detail below with reference to the figures.

<Basic Configuration of the Speech Synthesis Device>

Figure 1 shows the configuration of a speech synthesis device (text-to-speech conversion device) relating to an embodiment of this invention. Text containing a mixture of *kanji* and *kana* characters (referred to as a Japanese-language text document) is input to text analysis part 101, where this input document is subjected to morphological analysis by referring to a word dictionary 102, the pronunciation, accentuation and intonation of each morpheme obtained by this analysis are analyzed, and then phonological symbols (intermediate language) with associated prosodic symbols are output for each morpheme.

In parameter generation part 103, based on the intermediate language itself, the segment address to be used is selected from within a segment dictionary 105, and parameters such as the pitch frequency pattern, phoneme duration time and amplitude are set.

Segment dictionary 105 is produced beforehand by segment generation part 106 after inputting speech signals to segment generation part 106.

In segment generation part 106, before synthesizing speech, segments are produced beforehand from the speech data, on a base of which segments synthesized sound will be generated.

Waveform synthesis part 104 can apply various conventional methods as the waveform synthesis method; for example, it might use a pitch synchronous overlap add (PSOLA) method. Note that rule-based speech synthesis is the synthesis of speech from an input consisting of phonological symbols with associated prosodic symbols (intermediate language).

The phoneme duration time determined in parameter generation part 103 mainly regulates the phoneme duration time by extending or contracting the vowel parts based on the isochrony of the Japanese language. Specifically, processing is performed whereby either the tail end segment is used repeatedly (extension) when the determined phoneme duration time is longer than the segment, or is cut off mid-way (contraction) when the determined phoneme duration time is shorter.

Note that in Figure 1, text analysis part 101, word dictionary 102, waveform synthesis part 104, segment dictionary 105 and segment generation part 106 can be configured using conventional techniques.

<First Embodiment of Method for Setting the Phoneme Duration Time in the Parameter Generation Part>

A first embodiment of a method for setting the phoneme duration time in parameter generation part 103 will be described in detail with reference to Figure 2.

In Figure 2, a phoneme symbol sequence is input to a phoneme type judgement part 201, which judges whether the phoneme in question is a vowel or consonant and, in the case of a consonant, judges whether or not it is a consonant anteriorly having a closing interval (*/p, t, k/* etc.; see Fig. 6). As a result, it operates a vowel length estimation part 202 when it judges that the phoneme is a vowel, and when it judges that the phoneme is a consonant, it either operates a consonant length estimation part 205 or, when it has judged that this phoneme anteriorly has a closing interval (such as */p, t, k/*), it operates a closing length estimation part 208, whereby the respective time lengths are estimated. After that, the estimated time lengths are set by vowel length setting part 203, consonant length setting part 206 and closing length setting part 209, respectively. The consonant length setting is performed in the following temporal order: estimated closing length, followed by estimated consonant length. Note that as a result of our analyzing real speech data, it has been found that the types of consonants that anteriorly have a closing length are only the phonemes shown in Figure 6, and accordingly nasal and the like are not included.

Note that a Hayashi's first method of quantification can, for example, be used to estimate the temporal length. In this method, learning data 211 is used beforehand to learn each of the models in vowel length learning part 204, consonant length learning part 207 and closing length learning part 210 (corresponding to solving simultaneous equations on a basis such as the abovementioned equation (3)), and the weighting coefficients necessary for estimation are determined as a result of this learning. The weighting coefficient means x(jk) on the abovementioned equation (1).

As described above, the phoneme duration time setting method of the present embodiment makes it possible to control the appropriate phoneme duration time with respect to phonemes anteriorly having a closing interval, and accordingly it is possible to obtain a highly natural synthesized sound in a rule-based speech synthesis device.

Note that the present embodiment employs a configuration wherein a Hayashi's first method of quantification is used for learning and estimation, but is not limited thereto, and other statistical methods may also be used.

<Second Embodiment of Method for Setting the Phoneme Duration Time in the Parameter Generation Part>

A second embodiment of a method for setting the phoneme duration time in parameter generation part 103 will be described in detail with reference to Figure 3.

The configuration shown in Figure 3 differs from that of the first embodiment in that a closing length classification part 301 is provided, and in that closing length learning part 302 and closing length estimation part 303 operate differently; parts that operate in the same way as in the first embodiment are given the same numbers as in Figure 2. The operation of this embodiment is described below.

First, a phoneme symbol sequence is input to phoneme type judgement part 201, and this judgement part 201 judges whether the phoneme in question is a vowel or consonant and, in the case of a consonant, judges whether or not it is a consonant that anteriorly has a closing interval. As a result, it operates a vowel length estimation part 202 when it judges that the phoneme is a vowel, and when it judges that the phoneme is a consonant, it either operates a consonant length estimation part 205 or, when it has judged that this phoneme anteriorly has a closing interval, it operates a closing length estimation part 303, whereby the respective time lengths are estimated. After that, the estimated time lengths are set by vowel length setting part 203, consonant length setting part 206 and closing length setting part 209, respectively. The consonant length setting is performed in the following temporal order: estimated closing length, followed by estimated consonant length.

Hayashi’s first method of quantification is used to estimate the temporal length. However, in the second embodiment, the method whereby a Hayashi’s first method of quantification is used to learn/estimate the closing length differs from that of the first embodiment. Specifically, in Figure 3, learning data 211 is classified beforehand by a closing length classification part 301, each model of closing length learning part 302 is learned, and the weighting coefficients necessary for estimation are determined beforehand.

Since the Hayashi's first method of quantification performs modeling by a linear weighted sum of only the number of category numbers, the estimation precision is determined by the reliability of the learning data. Also, although the factors used in this modeling include the phoneme in question, the environment of the two phonemes before and after it and the position of the phoneme, these factors generally take the form of qualitative data and are not arranged in order of magnitude. Consequently, there is no way in which the factors can be essentially grouped.

In the second embodiment, closing length classification part 301, closing length learning part 302 and closing length estimation part 303 are provided to solve this problem and characterize this embodiment, and the operation thereof is described with reference to Figure 7.

In Figure 7, the frequency distribution of an external criterion (closing length) of the learning data is determined at step 701 in closing length classification part 301. At step 702, based on the frequency distribution, the closing lengths are divided into some groups. Furthermore, at step 703 the correspondence with the phoneme in question is obtained, and this phoneme is also divided into groups.

In closing length learning part 302, learning is performed for each of the abovementioned groups at step 704 and the weighting coefficients are determined, and as a result the weighting coefficients are transmitted to closing length estimation part 303 at step 705.

Next, estimation is performed. In closing length estimation part 303, the name of the phoneme in question is judged based on the input phoneme symbol sequence at step 710, said group is selected based on the name of the phoneme in question at step 711, the weighting coefficients inherent to said group are selected at step 712, and said weighting coefficients are used to estimate the closing length by a Hayashi's first method of quantification at step 713.

As described above, with the phoneme time length setting method of the present embodiment, by classifying the closing lengths into groups as described above, it is possible to obtain a desirable distribution of the closing lengths that actually appear. As a result, learning can be achieved with greater precision than in conventional methods and the distribution of estimated values can be kept small in the estimations, which has the advantage of improving the estimation precision.

<Third Embodiment of Method for Setting the Phoneme Duration Time in the Parameter Generation Part>

A third embodiment of a method for setting the phoneme duration time in parameter generation part 103 is described in detail with reference to Figure 4.

The configuration shown in Figure 4 differs from that of the second embodiment in that a vowel length classification part 401 and a consonant length classification part 404 are provided, and in that vowel length learning part 402, vowel length estimation part 403, consonant length learning part 405 and consonant length estimation part 406 operate differently; parts that operate in the same way as in the second embodiment are given the same numbers as in Figure 3. The operation of this embodiment is described below.

First, a phoneme symbol sequence is input to phoneme type judgement part 201, and this judgement part 201 judges whether the phoneme in question is a vowel or consonant and, in the case of a consonant, judges whether or not it is a consonant that anteriorly has a closing interval. As a result, it either operates vowel length estimation part 403 when it judges that the phoneme is a vowel, or it operates consonant length estimation part 406 when it judges that the phoneme is a consonant, or it operates closing length estimation part 303 when it judges that this phoneme anteriorly has a closing interval, whereby the respective time lengths are estimated. After that, the estimated time lengths are set respectively by vowel length setting part 203, consonant length setting part 206 and closing length setting part 209. The consonant length setting is performed in the following temporal order: estimated closing length, followed by estimated consonant length.

In Figure 4, the vowel length learning data in the previously learning data 211 is classified by a vowel length classification part 401, and the consonant length learning data is classified by a consonant length classification part 404. As for the closing length, the closing length learning data is classified by closing length classification part 301, and since closing length learning part 302 and closing length estimation part 303 are operated in the same way as in the second embodiment, their description is omitted here.

The factors of Hayashi's first method of quantification take the form of qualitative data and are not arranged in order of magnitude. Consequently, there is no way in which the factors can be essentially grouped. The third embodiment, like the second embodiment, aims to improve on this, and in particular it aims to improve the estimation precision of vowel length and consonant length.

The characterizing features of the third embodiment are vowel length classification part 401, vowel length learning part 402 and vowel length estimation part 403, whose operation is illustrated in Figure 8, and consonant length classification part 404, consonant length learning part 405 and consonant length estimation part 406, whose operation is illustrated in Figure 9.

In relation to the vowel length, the frequency distribution of an external criterion (vowel length) in the learning data is determined at step 801 in Figure 8. At step 802, based on the frequency distribution, the vowel length is divided into some groups. Furthermore, at step 803 the correspondence with the phoneme in question is obtained, and this phoneme is also divided into groups. In vowel length learning part 402, learning is performed for each of the abovementioned groups at step 804 and the weighting coefficients are determined, and as a result the weighting coefficients are transmitted to vowel length estimation part 403 at step 805.

When estimation is performed in vowel length estimation part 403, the name of the phoneme in question is judged from the input phoneme symbol sequence at step 810, said group is selected from the phoneme name in question at step 811, the weighting coefficients inherent to said group are selected at step 812, and said weighting coefficients are used to estimate the vowel length by Hayashi's first method of quantification at step 813.

Similarly, in relation to consonants, the frequency distribution of an external criterion (consonant length) in the learning data is determined at step 901 in Figure 9. At step 902, based on the frequency distribution, the consonant length is divided into some groups. Furthermore, at step 903 the correspondence with the phoneme in question is obtained, and this phoneme is also divided into groups. In consonant length learning part 405, learning is performed for each of the abovementioned groups at step 904 and the weighting coefficients are determined, and as a result the weighting coefficients are transmitted to consonant length estimation part 406 at step 905.

When estimation is performed in consonant length estimation part 406, the name of the phoneme in question is judged based on the input phoneme symbol sequence at step 910, said group is selected based on the phoneme name in question at step 911, the weighting coefficients inherent to said group are selected at step 912, and said weighting coefficients are used to estimate the consonant length by Hayashi's first method of quantification at step 913.

As described above, with the present embodiment, the vowel lengths and consonant lengths do not have simple distributions and generally have multi-peaked distributions. By classifying them into groups as described above, learning can be achieved with learning data that is more precise than in conventional methods and the distribution of estimated values can be kept small in the estimations, because the average values of the estimated values are the average values of said groups, thereby improving the estimation precision.

<Fourth Embodiment of Method for Setting the Phoneme Duration Time in the Parameter Generation Part>

A fourth embodiment of a method for setting the phoneme duration time in parameter generation part 103 will be described in detail with reference to Figure 5.

In Figure 5, blocks that function in the same way as those in Figure 2 and Figure 3 are given the same numbers. In Figure 5, closing length estimation part 208 comprises a factor extraction part 501, a prior de-voicing judgement means 502 and an estimation model part 503, and closing length learning part 210 consists of a factor extraction part 505, a prior de-voicing judgement means 506 and a learning model part 504. The operation of these parts will be described below.

First, the closing length learning data 510 in the learning data 211 is classified into groups by closing length classification part 303 in the same way as in the second embodiment. After that, factor extraction part 505 extracts factors such as the phoneme name in question, the environment of the two phonemes before and after it, the phoneme position (within a breath group, within a sentence), number of moras (breath group, sentence), part of speech and the like, quantizes these factors, and supplies the results to learning model part 504. At the same time, prior de-voicing judgement means 506 makes a judgement based on the learning data as to whether or not the previous phoneme is de-voiced. Numerical data with a value of 1 is generated if the result of this judgement is that the previous phoneme is to be de-voiced, while numerical data of a value of 2 is generated if it is judged not to be de-voiced, and this numerical data is supplied to learning model part 504. Learning model part 504 is configured to correspond to a model of Hayashi's first method of quantification. This model part 504 then produces a weighting coefficient table 520 for each factor as the learning results for each of said groups, and sends weighting coefficient table 520 to estimation model part 503.

During estimation, in factor extraction part 501, factors that are the same as those in factor extraction part 505 in closing length learning part 210 are extracted from the input phoneme symbol sequence, and these factors are quantized. At the same time, in prior de-voicing judgement means 502, de-voicing of the phoneme is judged by applying the de-voicing rules described below. Numerical data with a value of 1 is generated if the result of this judgement is that the phoneme prior to the phoneme in question is to be de-voiced, while numerical data with a value of 2 is generated if it is judged not to be de-voiced. In estimation model part 503, said group is judged from the phoneme in question, weighting coefficient table 520 is accessed for each group, and the closing length is estimated by a model of Hayashi's first method of quantification.

Here, the de-voicing rules include the following:

(1) An /*i*/ or /*u*/ sandwiched between unvoiced consonants is de-voiced.

However,

(2) De-voicing is not performed if the phoneme is accentuated.

(3) Consecutive de-voicing is not allowed.

(4) A vowel sandwiched between unvoiced fricatives of the same type is not de-voiced.

These rules are applied by analyzing the input phoneme symbol sequence.

As described above, with the present embodiment, since the closing length is controlled depending on whether or not the preceding phoneme is de-voiced, for example, since /*i*/ in the syllable /*chi*/ of /*ochikaku*/ ("nearby") is de-voiced, it is possible to control the closing interval length that prefixes the /*k*/ of the following syllable /*ka*/ to an appropriate value.

Although a configuration is employed wherein the de-voicing rules mentioned below are applied to determine the de-voicing of phonemes in the prior de-voicing judgement means 502 of the fourth embodiment, it is also possible – as an alternative embodiment – to employ a configuration wherein the application of de-voicing rules is performed separately beforehand and predetermined de-voicing information is obtained in closing length estimation part 208.

As described in detail above, since the present invention is a rule-based speech synthesis device that generates arbitrary speech by selecting and concatenating previously stored speech synthesis units and controlling the prosodic information and which is configured by providing it with a phoneme duration time setting means that estimates and controls the closing interval length of phonemes having a closing interval separately for the vowel length and consonant length, it is possible to control the suitable phoneme duration time for phonemes anteriorly having a closing interval, and it is possible to obtain very natural-sounding synthesized speech from a rule-based speech synthesis device.

What is claimed is:

1. A rule-based speech synthesis device which synthesizes arbitrary speech by selecting and concatenating previously stored speech synthesis units and controlling prosodic information, comprising:

a phoneme duration time setting means which estimates and controls the closing interval length of a phoneme having a closing interval, independently of the vowel length and consonant length.

2. The rule-based speech synthesis device according to claim 1, wherein said phoneme duration time setting means comprises:

a phoneme type judgement means that judges the type of a phoneme with respect to the input phoneme symbol sequence,

a vowel length determining means comprising a vowel length estimation means and a vowel length learning means,

a consonant length determining means comprising a consonant length estimation means and a consonant length learning means, and

a closing length determining means comprising a closing length estimation means and a closing length learning means;

and wherein said phoneme type judgement means operates said vowel length estimation means or consonant length estimation means depending on whether the phoneme in question is a vowel or a consonant, and if it is judged to be a consonant, it judges whether or not it anteriorly has a closing interval and if it anteriorly has a closing interval then it operates a closing length estimation means.

3. The rule-based speech synthesis device according to claim 2, wherein:

said closing length determining means further comprises a closing length classification means;

said closing length classification means performs classification operations whereby it obtains a frequency distribution of closing lengths from learning data, classifies the closing lengths into a first group based on said frequency distribution and classifies the phoneme in question into a second group based on the first group;

said closing length learning means performs learning operations whereby it is learned with each member of the said second group and outputs weighting coefficients - which are necessary for estimation of phoneme duration times - to the closing length estimation means; and

said closing length estimation means judges the name of the phoneme in question from an input phoneme symbol sequence, judges and selects said second group from said phoneme name, selects weighting coefficients inherent to said group, performs operations to estimate the closing length using said weighting coefficients, and outputs the value of the estimated closing length.

4. The rule-based speech synthesis device according to claim 2, wherein:

said vowel length determining means further comprises a vowel length classification means;

said vowel length classification means performs classification operations whereby it obtains a frequency distribution of vowel lengths from learning data, classifies the vowel lengths into a first group based on said frequency distribution and classifies the phoneme in question into a second group based on the first group;

said vowel length learning means performs learning operations whereby it is learned with each member of the said second group and outputs weighting coefficients - which are necessary for estimation of phoneme duration times - to the vowel length estimation means; and

said vowel length estimation means judges the name of the phoneme in question from an input phoneme symbol sequence, judges and selects said second group from said phoneme name, selects weighting coefficients inherent to said group, performs operations to estimate the vowel length using said weighting coefficients, and outputs the value of the estimated vowel length.

5. The rule-based speech synthesis device according to claim 2, wherein:

said consonant length determining means further comprises a consonant length classification means;

said consonant length classification means performs classification operations whereby it obtains a frequency distribution of consonant lengths from learning data, classifies the consonant lengths into a first group based on said frequency distribution and classifies the phoneme in question into a second group based on the first group;

said consonant length learning means performs learning operations whereby it is learned with each member of the said second group and outputs weighting coefficients - which are necessary for estimation of phoneme duration times - to the consonant length estimation means; and

said consonant length estimation means judges the name of the phoneme in question from an input phoneme symbol sequence, judges and selects said second group from said phoneme name, selects weighting coefficients inherent to said group, performs operations to estimate the consonant length using said weighting coefficients, and outputs the value of the estimated consonant length.

6. The rule-based speech synthesis device according to claim 3, wherein:

said closing length learning means is composed of a first factor extraction means which extracts and quantizes factors comprising the phoneme in question, the phoneme environment consisting of the two phonemes before and after the phoneme in question, the phoneme position, the part of speech and the like, a first prior de-voicing judgement means which judges whether or not the previous phoneme is de-voiced based on the learning data, and a model learning means which produces weighting coefficients for each factor in each of said classified second groups;

and wherein said closing length estimation means is composed of a second factor extraction means which extracts and quantizes factors comprising the phoneme in question, the phoneme environment consisting of the two phonemes before and after the phoneme in question, the phoneme position, the part of speech and the like, a second prior de-voicing judgement means which judges whether or not the phoneme in question is to be de-voiced based on prescribed de-voicing rules, and a model estimation means which judges said second group from the phoneme in question and estimates the closing length by referring to the weighting coefficients output from said model learning means for each group.

ABSTRACT OF THE DISCLOSURE

The principal object of this invention is to provide a suitable control method for closing length with respect to phonemes (such as unvoiced plosive consonants) having a closing interval, and as a result an improved rule-based speech synthesis device is provided. A phoneme type judgement part 201 judges whether the phoneme in question is a vowel or consonant and, in the case of a consonant, judges whether or not it is a consonant that anteriorly has a closing interval. As a result, it operates a vowel length estimation part 202 when it judges that the phoneme is a vowel and operates a consonant length estimation part 205 when it judges that the phoneme is a consonant, and when it has judged that this phoneme anteriorly has a closing interval, it operates a closing length estimation part 208, whereby the respective time lengths are estimated. After that, the estimated time lengths are set by vowel length setting part 203, consonant length setting part 206 and closing length setting part 209, respectively.

FIG. 1

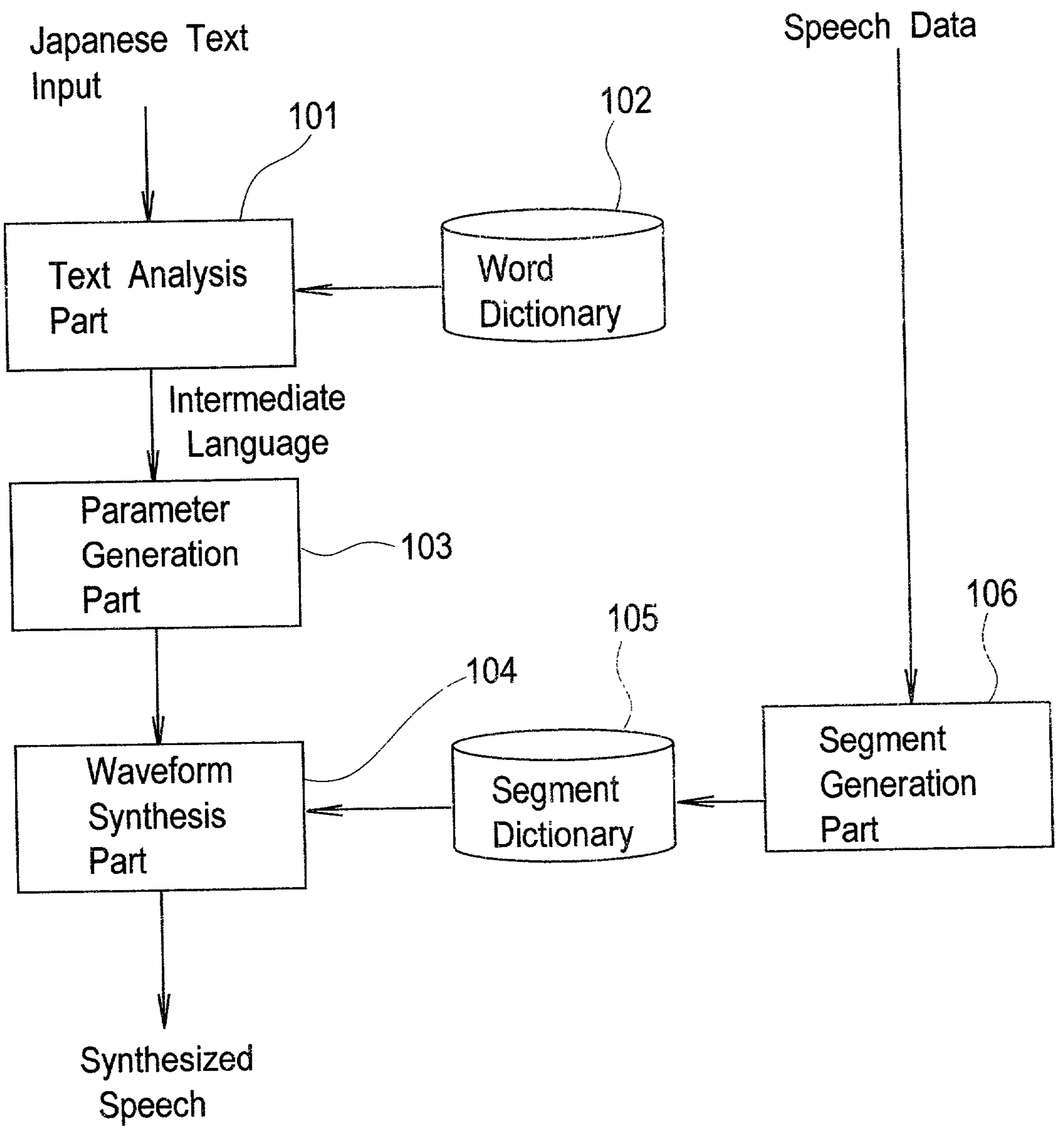

Japanese Text Input
101
102
Text Analysis Part
Word Dictionary
Intermediate Language
Parameter Generation Part
103
Speech Data
106
105
104
Waveform Synthesis Part
Segment Dictionary
Segment Generation Part
Synthesized Speech

FIG.2

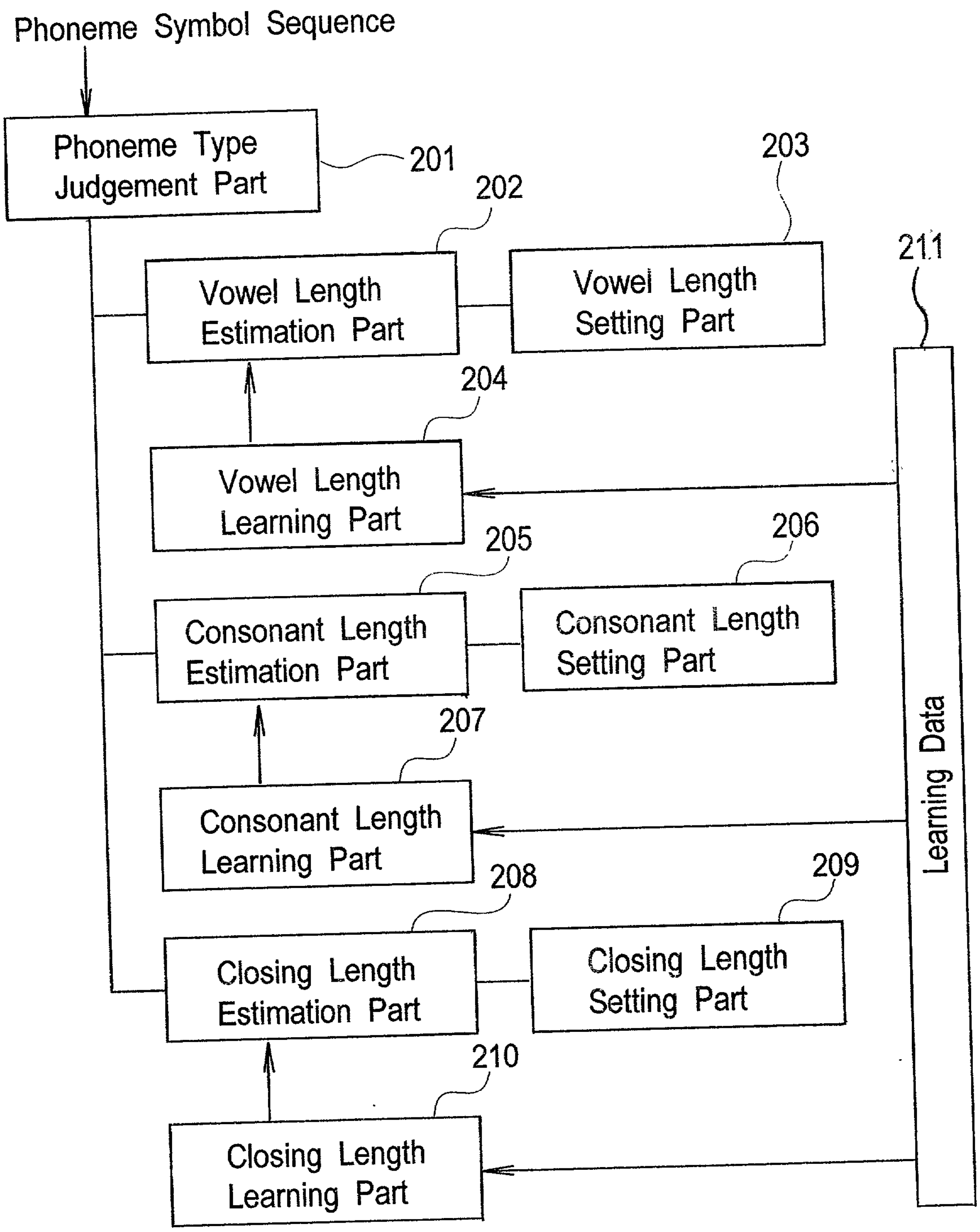

Phoneme Symbol Sequence
Phoneme Type Judgement Part
201
202
203
Vowel Length Estimation Part
Vowel Length Setting Part
211
204
Vowel Length Learning Part
205
206
Consonant Length Estimation Part
Consonant Length Setting Part
207
Consonant Length Learning Part
208
209
Closing Length Estimation Part
Closing Length Setting Part
210
Closing Length Learning Part
Learning Data

FIG.3

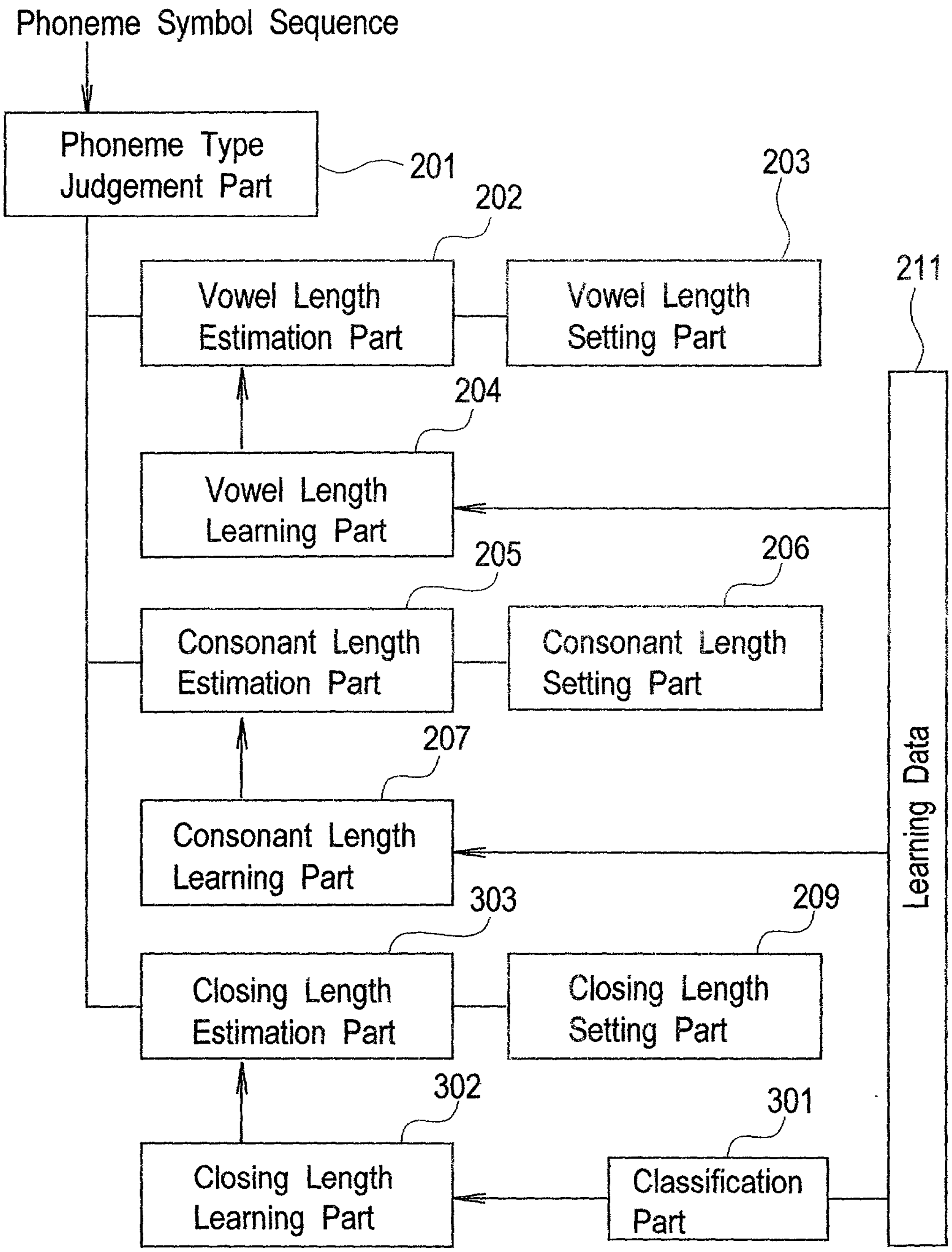

Phoneme Symbol Sequence
Phoneme Type Judgement Part
201
202
203
Vowel Length Estimation Part
Vowel Length Setting Part
211
204
Vowel Length Learning Part
205
206
Consonant Length Estimation Part
Consonant Length Setting Part
207
Consonant Length Learning Part
Learning Data
303
209
Closing Length Estimation Part
Closing Length Setting Part
302
301
Closing Length Learning Part
Classification Part

FIG. 4

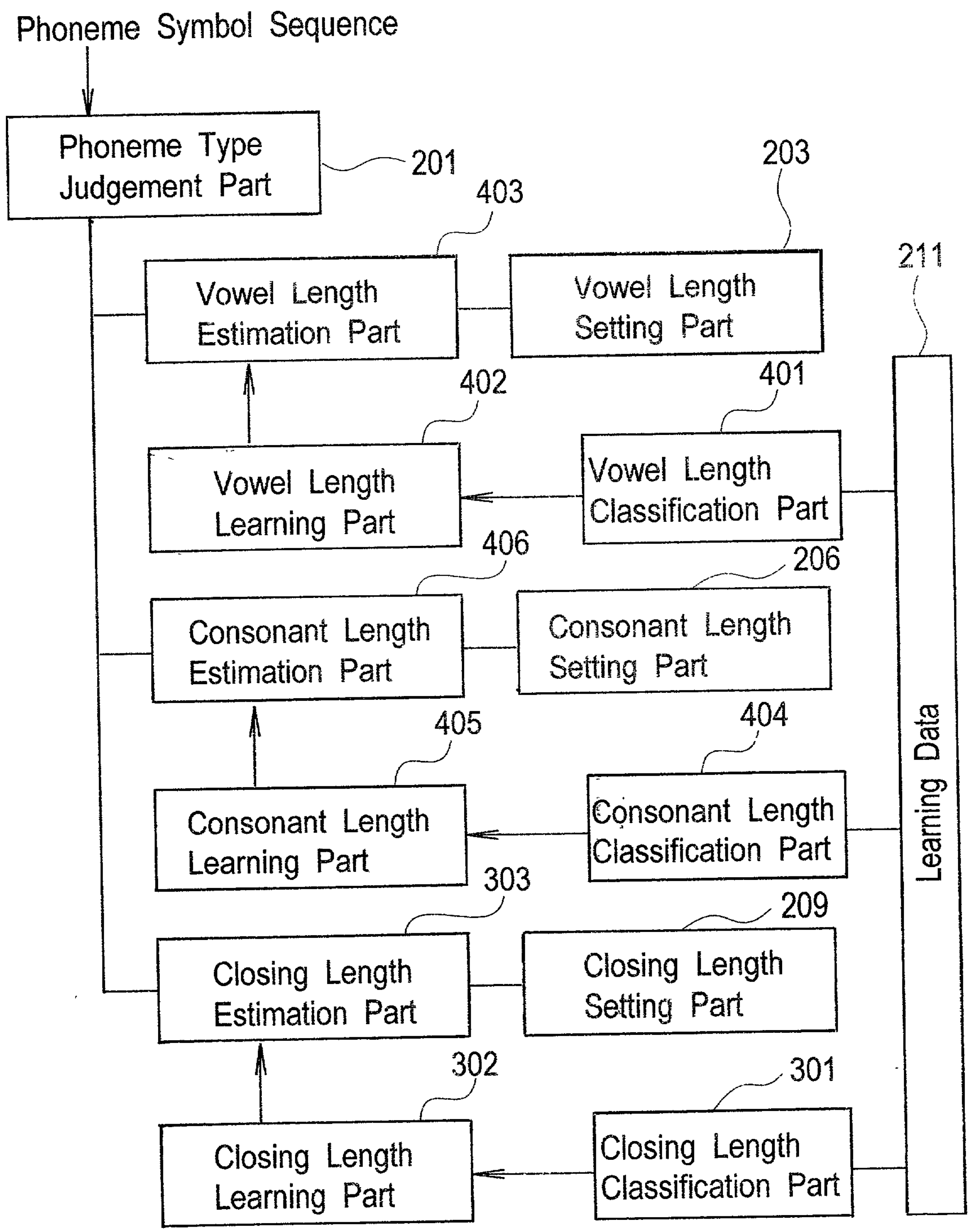

Phoneme Symbol Sequence
Phoneme Type Judgement Part
201
203
403
Vowel Length Estimation Part
Vowel Length Setting Part
211
402
401
Vowel Length Learning Part
Vowel Length Classification Part
406
206
Consonant Length Estimation Part
Consonant Length Setting Part
405
404
Consonant Length Learning Part
Consonant Length Classification Part
Learning Data
303
209
Closing Length Estimation Part
Closing Length Setting Part
302
301
Closing Length Learning Part
Closing Length Classification Part

FIG.5

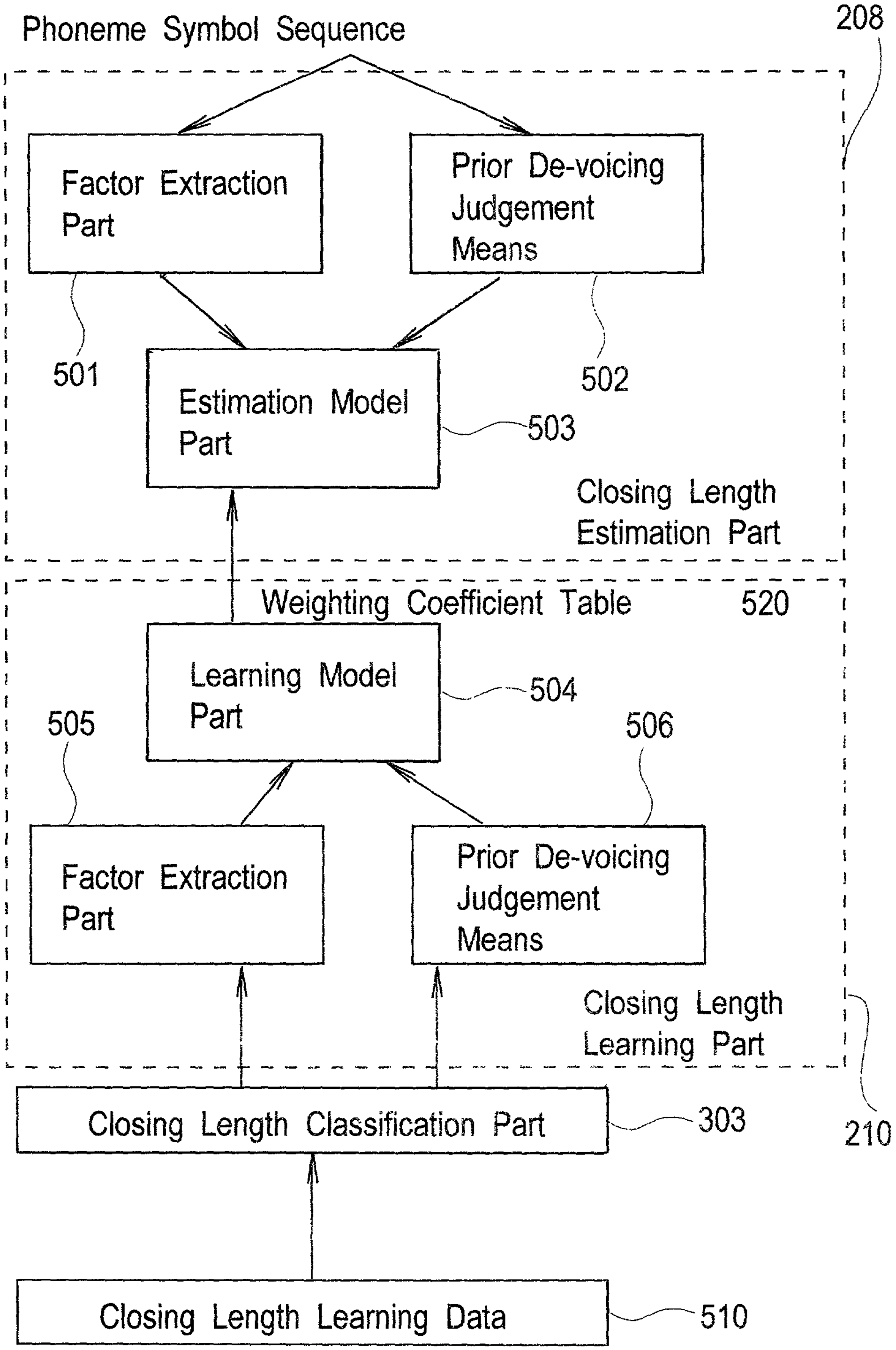
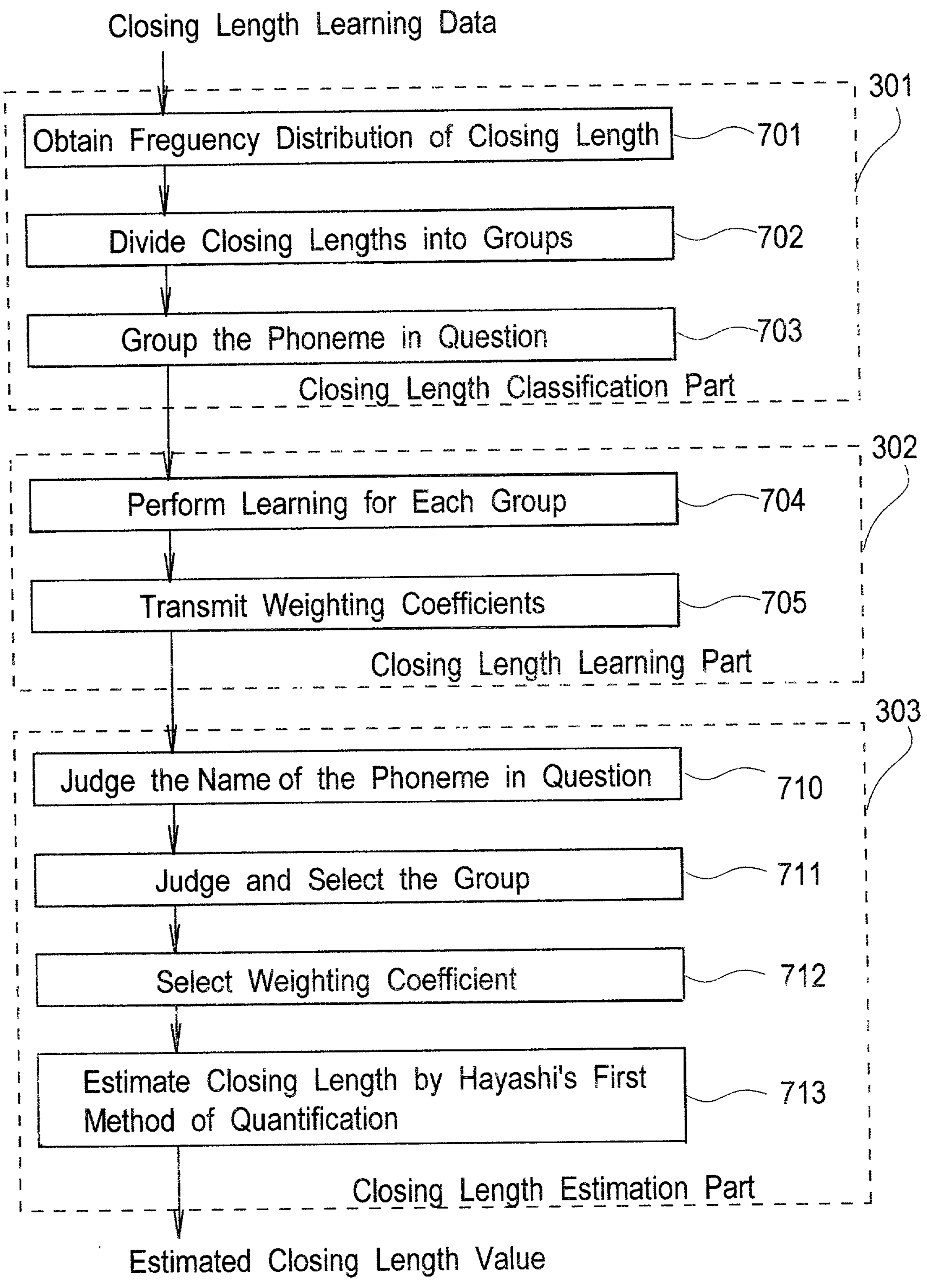

Phoneme Symbol Sequence
208
Factor Extraction Part
Prior De-voicing Judgement Means
501
502
Estimation Model Part
503
Closing Length Estimation Part
Weighting Coefficient Table
520
Learning Model Part
504
505
506
Factor Extraction Part
Prior De-voicing Judgement Means
Closing Length Learning Part
Closing Length Classification Part
303
210
Closing Length Learning Data
510

FIG.6

Number	Type Name	Phoneme Symbol
1	Unvoiced Plosive Consonant	/ p,t,k /
2	Some Palatalized Sound	/ py,ky,cy, /
3	Others	/ ts,ch /

FIG. 7



Closing Length Learning Data
301
Obtain Freguency Distribution of Closing Length
701
Divide Closing Lengths into Groups
702
Group the Phoneme in Question
703
Closing Length Classification Part
302
Perform Learning for Each Group
704
Transmit Weighting Coefficients
705
Closing Length Learning Part
303
Judge the Name of the Phoneme in Question
710
Judge and Select the Group
711
Select Weighting Coefficient
712
Estimate Closing Length by Hayashi's First Method of Quantification
713
Closing Length Estimation Part
Estimated Closing Length Value

FIG.8

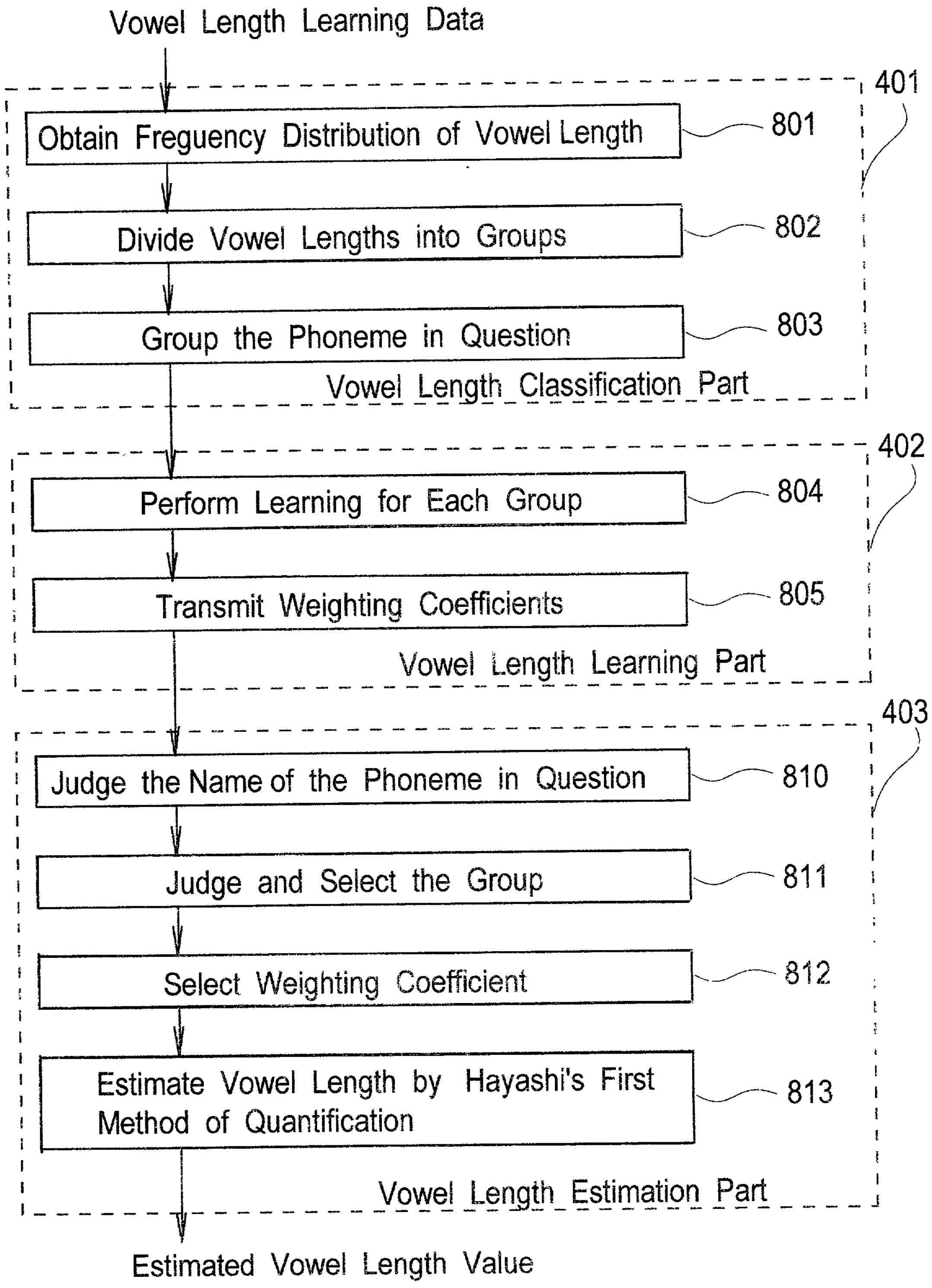
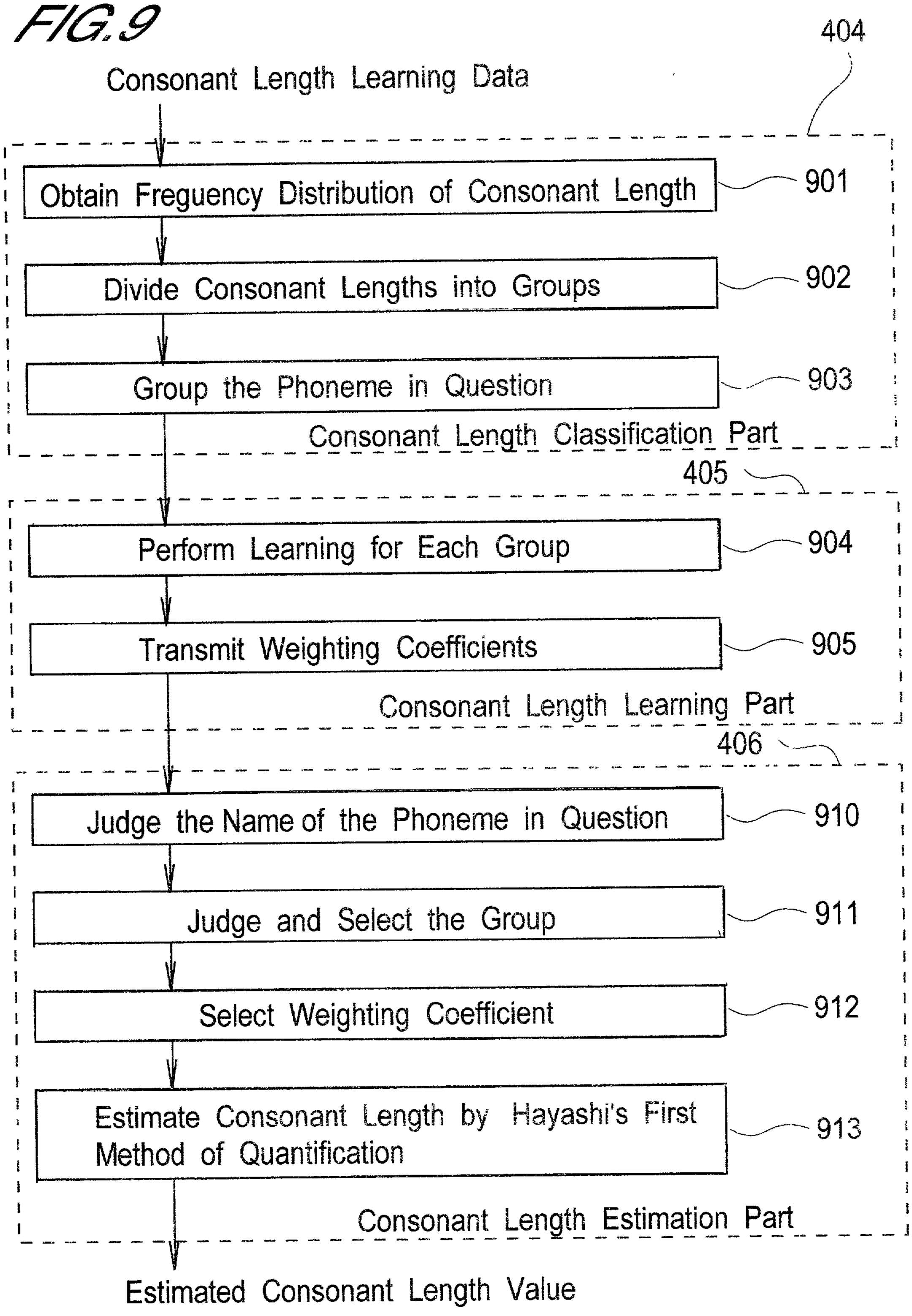

Vowel Length Learning Data
401
Obtain Freguency Distribution of Vowel Length
801
Divide Vowel Lengths into Groups
802
Group the Phoneme in Question
803
Vowel Length Classification Part
402
Perform Learning for Each Group
804
Transmit Weighting Coefficients
805
Vowel Length Learning Part
403
Judge the Name of the Phoneme in Question
810
Judge and Select the Group
811
Select Weighting Coefficient
812
Estimate Vowel Length by Hayashi's First Method of Quantification
813
Vowel Length Estimation Part
Estimated Vowel Length Value

FIG. 9


404
Consonant Length Learning Data
Obtain Freguency Distribution of Consonant Length
901
Divide Consonant Lengths into Groups
902
Group the Phoneme in Question
903
Consonant Length Classification Part
405
Perform Learning for Each Group
904
Transmit Weighting Coefficients
905
Consonant Length Learning Part
406
Judge the Name of the Phoneme in Question
910
Judge and Select the Group
911
Select Weighting Coefficient
912
Estimate Consonant Length by Hayashi's First Method of Quantification
913
Consonant Length Estimation Part
Estimated Consonant Length Value

PTO/SB/106(8-96)
Approved for use through 9/30/98. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Declaration and Power of Attorney For Patent Application

特許出願宣言書及び委任状

Japanese Language Declaration

日本語宣言書

下記の氏名の発明者として、私は以下の通り宣言します。

As a below named inventor, I hereby declare that:

私の住所、私書箱、国籍は下記の私の氏名の後に記載された通りです。

My residence, post office address and citizenship are as stated next to my name.

下記の名称の発明に関して請求範囲に記載され、特許出願している発明内容について、私が最初かつ唯一の発明者（下記の氏名が一つの場合）もしくは最初かつ共同発明者であると（下記の名称が複数の場合）信じています。

____

____

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

SPEECH SYNTHESIS DEVICE

____

上記発明の明細書（下記の欄で×印がついていない場合は、本書に添付）は、

the specification of which is attached hereto unless the following box is checked:

□ ＿月＿日に提出され、米国出願番号または特許協定条約国際出願番号を＿＿＿＿＿とし、
（該当する場合）＿＿＿＿＿に訂正されました。

□ was filed on ______________________
as United States Application Number or
PCT International Application Number
______________ and was amended on
______________ (if applicable).

私は、特許請求範囲を含む上記訂正後の明細書を検討し、内容を理解していることをここに表明します。

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

私は, 連邦規則法典第３７編第１条５６項に定義されるとおり、特許資格の有無について重要な情報を開示する義務があることを認めます。

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

Burden Hour Statement: This form is estimated to take 0.4 hours to copulate, Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington,DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner of Patents and Trademarks, Washington, DC 20231.

Japanese Language Declaration
(日本語宣言書)

私は、米国法典第３５編１１９条（a）－（d）項又は３６５条（b）項に基き下記の、米国以外の国の少なくとも一ヵ国を指定している特許協力条約３６５（a）項に基づく国際出願、又は外国での特許出願もしくは発明者証の出願についての外国優先権をここに主張するとともに、優先権を主張している、本出願の前に出願された特許または発明者証の外国出願を以下に、枠内をマークすることで、示しています。

I hereby claim foreign priority under Title 35, United States Code. Section 119 (a)-(d) or 365 (b) of any foreign application(s) for patent or inventor's certificate, or 365 (a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent of inventor's certificate, or PCT International application having a filing date before that of the application on which priority is clamed.

Prior Foreign Application(s)
外国での先行出願

(Number) (番号)	(Country) (国名)	(Day/Month/Year Filed) (出願年月日)	Priority Not Claimed 優先権主張なし
075831/2000	JAPAN	March 17, 2000	□
			□

私は、第３５編米国法典１１９条（e）項に基いて下記の米国特許出願規定に記載された権利をここに主張いたします。

I hereby claim the benefit under Title 35, United States Code, Section 119(e) of any United States provisional application(s) listed below.

(Application No.) (出願番号)   (Filing Date) (出願日)

(Application No.) (出願番号)   (Filing Date) (出願日)

私は、下記の米国法典第３５編１２０条に基いて下記の米国特許出願に記載された権利、又は米国を指定している特許協力条約３６５条（c）に基づく権利をここに主張します。また、本出願の各請求範囲の内容が米国法典第３５編１１２条第１項又は特許協力条約で規定された方法で先行する米国特許出願に開示されていない限り、その先行米国出願書提出日以降で本出願書の日本国内または特許協力条約国際提出日までの期間中に入手された、連邦規則法典第３７編１条５６項で定義された特許資格の有無に関する重要な情報について開示義務があることを認識しています。

I hereby claim benefit under Title 35, United States Code, Section 120 of any United States application(s), or 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States of PCT International application in the manner provided by the first paragraph of Title 35, United States Code Section 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of application.

(Application No.) (出願番号)   (Filing Date) (出願日)   (Status: Patented, Pending, Abandoned) (現況： 特許許可済、係属中、放棄済)

(Application No.) (出願番号)   (Filing Date) (出願日)   (Status: Patented, Pending, Abandoned) (現況： 特許許可済、係属中、放棄済)

私は、私自身の知識に基づいて本宣言書中で私が行なう表明が真実であり、かつ私の入手した情報と私の信じるところに基づく表明が全て真実であると信じていること、さらに故意になされた虚偽の表明及びそれと同等の行為は米国法典第１８編第１００１条に基づき、罰金または拘禁、もしくはその両方により処罰されること、そしてそのような故意による虚偽の声明を行なえば、出願した、又は既に許可された特許の有効性が失われることを認識し、よってここに上記のごとく宣誓を致します。

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and the such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Japanese Language Declaration
（日本語宣言書）

委任状：　私は下記の発明者として、本出願に関する一切の手続きを米特許商標局に対して遂行する弁理士または代理人として、下記の者を指名いたします。（弁護士、または代理人の氏名及び登録番号を明記のこと）

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith *(list name and registration number)*

George H. Spencer (Reg. No. 18,038), **Robert J. Frank** (Reg. No. 19,112), **Norman N. Kunitz** (Reg. No. 20,586), **Gabor J. Kelemen** (Reg. No. 21,016), **John W. Schneller** (Reg. No. 26,031), **Marina V. Schneller** (Reg. No. 26,032) **Robert Kinberg** (Reg. No. 26,924), **Allen Wood** (Reg. No. 28,134), **Ashley J. Wells** (Reg. No. 29,847), **Richard d. Schmidt** (Reg . No. 31.301), **James R. Burdett** (Reg. No. 31,594), **Michael A. Gollin** (Reg. No. 31,957), **Leo J. Jennings** (Reg. No. 32,902), **Catherine M. Voorhees** (Reg. No 33,074), **Gary L. Shaffer** (Reg. No. 34,502), **G. Abe Zachariah** (Reg. No. 38,366), **Julie A. Petruzzelli** (Reg. No. 40,769), **Catherine A. Ferguson** (Reg. No. 40,877), **Michael A. Sartori** (Reg. No. 41,289), **Fei-Fei Chao** (Reg. No. 43,538), **Charles C. P.Rories** (Reg. No. 43,381) and **Jeffrey W. Gluck** (Reg. No. 44,457), all at 1201 New York Avenue, N. W., Suite 1000, Washington, D.C. 20005-3917.

書類送付先　　Send Correspondence to:

Address all correspondence to VENABLE, Post Office Box 34385, Washington, D.C. 20043-9998.

直接電話連絡先：　（名前及び電話番号）　　Direct Telephone Calls to: *(name and telephone number)*

Robert J. Frank
VENABLE
Telephone: (202) 962-4800, Telefax: (202) 962-8300

唯一または第一発明者名		Full name of sole or first inventor Yukio TABEI	
発明者の署名	日付	Inventor's signature Yukio Tabei	Date October 16, 2000
住所		Residence Tokyo, Japan	
国籍 Japanese		Citizenship Japanese	
私書箱		Post Office Address c/o Oki Electric Industry Co., Ltd.	
		7-12, Toranomon 1-chome, Minato-ku, Tokyo, Japan	
第二共同発明者名		Full name of second joint inventor, if any	
第二共同発明者の署名	日付	Second inventor's signature	Date
住所		Residence	
国籍		Citizenship	
私書箱		Post Office Address	

（第三以降の共同発明者についても同様に記載し、署名をすること）

(Supply similar information and signature for third and subsequent joint inventors.)